DOCUMENT RESUME

ED 174 034 FL 010 490

AUTHOR Bond, Z. S.
TITLE Listening to Elliptic Speech: Pay Attention to Stressed Vowels.
PUB DATE [79]
NOTE 22p.; Expanded version of paper presented at annual meeting of the Linguistic Society of America (Boston, Massachusetts, December 28-30, 1978)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Acoustic Phonetics; Artificial Speech; *Auditory Perception; College Students; Language Research; Listening Comprehension; Memory; Perception Tests; Phonemes; *Phonology; *Pronunciation; Speech Communication; Speech Evaluation; *Suprasegmentals; *Vowels
IDENTIFIERS *Elliptic Speech

ABSTRACT

University students were the subjects of three experiments designed to determine the usefulness of elliptic speech in investigating the perception of the phonological structure of continuous speech. Five naturally spoken and five synthesized paragraphs were recorded in two different randomizations of phonological distortions and at two different signal to noise ratios. All three experiments tested the comprehensibility of the paragraphs. The synthesized paragraphs were found to be so nearly incomprehensible that they were effectively excluded from the results. Data indicated that: (1) comprehensibility decrement under phonological distortions is consistently lower for nasality, voicing, and place than for vowels; (2) although veridical reconstruction decreased, subjects found the speech more comprehensible in the presence of masking noise than when the signal was unimpeded by noise, indicating that a less available acoustic signal enables subjects to hypothesize a lexical match from other sources of knowledge; (3) the consistency of the results makes it reasonable to infer that responses to elliptic speech can be employed to assess continuous speech processing; and (4) a heuristic strategy employed in the perception of continuous speech seems to be "Pay attention to stressed vowels." Data are presented in graphic and tabular form. (JB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174034

LISTENING TO ELLIPTIC SPEECH: PAY ATTENTION TO STRESSED VOWELS

Z. S. Bond

School of Hearing and Speech Sciences

Ohio University, Athens, OH

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Z.S. Bond

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL010490

In their classic paper 'An Analysis of Perceptual Confusions Among Some English Consonants,' Miller and Nicely (1955) introduce 'elliptic' speech. Elliptic speech, or elliptic English, is constructed by replacing all consonants confused with one another under particular listening conditions by one consonant.

> Over a high quality communication system the elliptic speech was intelligible but sounded a little as though the talker had a marked dialect or speech defect. Then the low-pass filters were introduced ... (the conditions for which the substitutions were designed); the ellipsis could no longer be detected. Elliptic speech sounded just the same as normal speech under these conditions of distortion. (Miller and Nicely, p. 315).

Miller and Nicely were employing elliptic speech primarily in order to validate their description of English consonants as composed of a set of phonetic properties or features. However, elliptic speech, or speech very closely akin to it, might also be useful for another purpose: investigating the perception of the phonological structure of continuous speech.

If the phonological structure of an utterance were to be distorted by means of systematic phonological substitutions, then altering those properties which are perceptually most salient would be expected to lead to maximal disruption of the perception and/or comprehension of the utterance.[1] Thus, the relative degree of disruption of comprehension, under a particular phonological substitution

condition, could be taken as indicative of the relative perceptual salience of the distorted phonological property of the utterance.

The three experiments which will be described here represent an attempt to validate the elliptic speech technique as well as to provide an initial assessment of the relative perceptual saliency of some phonological properties.

All three of the experiments employed the same elliptic speech texts; consequently, the construction of the texts will be described first.

Five paragraphs of descriptive prose were selected from Arthur Clarke's science fiction novel Rendezvous with Rama (1973). Four paragraphs were subjected to four different types of systematic phonological distortion, conveniently labeled voicing, nasality, place, and vowels.

For voicing distortion, each voiced obstruent in the text was replaced by a voiceless obstruent, and vice versa; for nasality distortion, each nasal stop consonant was replaced by a voiced oral stop consonant, and vice versa; for place distortion, each stop and nasal was replaced by a stop and nasal produced at a different point of articulation, the replacing point of articulation being selected at random; and for vowel distortion, each stressed front vowel was replaced by the corresponding back vowel,

and vice versa. If a replacement would have violated English phonotactic constraints, it was not made. The paragraphs ranged in length from 61 to 71 words and were judged to be of equal difficulty.

The undistorted paragraph and the four phonologically distorted paragraphs were recorded by an experienced and highly-practiced male speaker in a natural reading style. Recordings of the same paragraphs were also made as produced by a Rockland digital speech synthesizer, controlled by a synthesis-by-rule program on a PDP-11 computer (Warmouth, 1975).

The five naturally spoken paragraphs and the five synthesized paragraphs were recorded in two different randomizations and at two different signal to noise ratios: with no added noise and mixed with white noise at a signal to noise ratio of +5 dB.

EXPERIMENT I

The purpose of the first experiment was to obtain an impressionistic estimate of the intelligibility of the four phonologically distorted paragraphs.

Subjects: The subjects were nine graduate students in the School of Hearing and Speech Sciences at Ohio University. The subjects performed the experimental task as part of a course requirement.

Procedure: The subjects listened to the four phonologically-distorted spoken paragraphs as recorded, and in the presence of masking noise. They were instructed to rank each paragraph on a scale ranging from 1, defined as completely unintelligible, to 7, defined as perfectly intelligible.

Results: The average rating for each conditions for the nine subjects is presented in Table I. The subjects judged the paragraphs presented under noise somewhat more intelligible than the recordings with no added noise, a result anticipated by the Miller and Nicely (1955) study. In addition, the subjects judged the vowel distortion condition to be by far the most unintelligible, with a mean rating of 2.0 and 2.4 in the two conditions. The voicing distortion condition was judged the least disruptive, with a mean rating of 5.0 and 5.8, respectively.

EXPERIMENT II

The second experiment was designed to assess the effect of phonological distortion on comprehension employing a response task dependent on memory.

Stimuli: All ten paragraphs, 5 naturally spoken and five synthetic, at two different signal-to-noise ratios, were used in this experiment.

Subjects: The subjects were native English speakers. All were students at Ohio University who volunteered to prticipate in the experiment. Each subject was screened for

hearing within normal limits (ANSI, 1969) and reported no speech/language difficulty. Forty subjects participated in the experiment, 10 in each listening condition.

Procedure: Subjects listened to one of the four listening tapes--natural speech and synthetic speech, with and without masking noise--over earphones at a comfortable listening level, individually or in small groups. After they heard each paragraph, subjects were given an opportunity to write down as much of it as they could remember. Subjects were encouraged to guess if they were not sure what particular words were.

The writen records of what each subject could remember from the paragraphs were scored in terms of the percent correct recall of predetermined content words. Function words and syntactic structure were not counted.

Results: In general, subjects found the task to be quite difficult, both because of listening conditions and memory load.

The synthetic speech paragraphs were all but unintelligible to the subjects, who reported less than five per cent of the words correctly, even without masking noise. Performance was so poor that any detailed evaluation of the response patterns appears to be meaningless. It is worth noting, however, that vowel distortion in synthetic speech was the experimental condition providing the most

difficulty: no content words whatever were reported correctly by any of the subjects in this condition.

Subjects found the natural speech paragraphs somewhat more manageable, reporting 11 to 14% of the content words correctly. The relative difficulty of the paragraphs is presented graphically in Fig. 1, listening without masking noise, and Fig. 2, listening with masking noise. In the figures, the per cent correct report for undistorted speech is taken as a base line. The figures show the percent decrement in performance when listening under each phonological distortion condition. The most disruption in recall was caused by distortion of stressed vowels, followed by distortion of place of articulation. The rank order of difficulty of the four distorted paragraphs was the same for both listening conditions.

In addition to examining the over-all difficulty with recall of the paragraphs, the subjects' written responses were analyzed for the number of veridical reports of the distorted words in comparison with the number of reports of another word, i.e. a restoration. These data are reported in Table II. When listening to natural speech without masking noise, the ratio of veridical report to restoration is approximately 1 to 2; in the presence of masking noise, this ratio becomes approximately 1 to 3. Apparently, masking noise, by making the acoustic signal less available, enables subjects to hypothesize a lexical match more readily

from other sources of knowledge. Perhaps the judgement that paragraphs are more intelligible in the presence of noise is a result of this effect.

EXPERIMENT III

The purpose of the third experiment was to assess the relative decrement in comprehension of the distorted paragraphs under conditions which by-passed memory limitations as much as possible. The response task selected was shadowing.

Stimuli: The five natural speech paragraphs, presented with and without masking noise, were employed in this experiment. Since the synthetic speech paragraphs were found to be extremely difficult, they were not employed.

Subjects: Thirty students at Ohio University volunteered to participate in the experiment. All subjects were screened for normal hearing (ANSI, 1969) and reported no speech/language difficulties.

Procedure: Each subject listened to only one paragraph in one listening condition. He was instructed to repeat what he heard on the stimulus tape, i.e. to shadow the paragraph. The subjects' responses were recorded on tape and scored as in the previous experiment.

Results: Subjects found the shadowing task much easier than the memory task. The pattern of relative difficulty of the paragraphs, however, was much the same as before. Fig. 3 presents the per cent decrement in shadowing performance in

listening to the paragraphs without noise; Fig. 4 presents the relative decrement in performance when listening to the paragraphs with masking noise present.

The rank order of difficulty of the distorting conditions is about the same as found previously: the greatest decrement in performance is produced by substitutions of stressed vowels, followed by substitutions of the place of articulation of stops and nasals.

Transcripts of the subjects' attempts at shadowing were analyzed to compare the proportion of veridical reports with restorations (See Table II). Shadowing without noise lead to a ratio of veridical report to restoration of approximately 1 to 2; with masking noise present, the ratio increased to approximately 1 to 5. As in the previous experiment, in the presence of noise when the acoustic signal is less readily available, subjects seem to rely considerably more on other sources of knowledge for achieving a hypothesized lexical match.

DISCUSSION

The purpose of the experiments was two-fold: to assess the usefulness of elliptic speech as a tool for investigating the perceptual salience of various phonological properties of connected speech and to make some preliminary inferences concerning the relative salience of

the properties investigated.

In all the conditions employed in the experiments, synthetic speech vs. natural speech, presence vs. absence of masking noise, written report vs. shadowing, the rank order of difficulty of the phonologicaldistortion conditions was similar.

Because of the consistency of the results, it is reasonable to infer that responses to elliptic speech can be employed to assess continuous speech processing in a useful manner.

The experiments reported here employed a rather simple phonetic classification, i.e. nasality, place and voicing, and the front/back vowel dimension. The overwhelming difficulties encountered with vowel distortion, in comparison with the other conditions, imply that the contribution of stressed vowels to the perception of continuous speech is very great.

This result confirms an observation based on the properties of misperceptions observed in spontaneous conversation. My colleague Sara Garnes and I (Garnes and Bond, 1978) have found that errors in the perception of stressed vowels are exceedingly rare in conversation; in a corpus of over 1000 misperceptions, we have found that only 4.1% of the errors involve misidentification of stressed vowels. From acoustical considerations, it is not

surprising that stressed vowel errors are rare, since the acoustic structure of vowels is characterized by greater intensity and duration than that of consonants. It is to be expected that listeners take advantage of the relative clarity and distinctness with which the phonetic information concerning vowel identity is made available in the speech signal. When, as in these experiments, the vowel information is distorted, i.e. when subjects are hearing incorrectly produced vowels, then the decrement in performance would be expected to be correspondingly great.

It appears, therefore, that on heuristic strategy employed in the perception of continuous speech is: 'Pay attention to stressed vowels.'

ACKNOWLEDGEMENTS

I am indebted to many people for the successful completion of this study. I want to thank: Donna Demerjian, who ran many of the experimental sessions and performed data analysis; James F. Naas, who served as a most talented speaker; Donald Warmuth, who produced the synthetic speech; and Robert Watkins, who provided many useful suggestions.

Footnote

[1] A conceptually related attempt to investigate syntactic processing in language use has been employed by by Shih, Carden, and Lane (1977), who employed various syntactically deviant sentences in an interview situation, using subjects' appropriate answers as the criterion of comprehension.

A preliminary version of this paper was presented at the annual meeting of the Linguistic Society of America, Boston, December 28-30, 1978.

REFERENCES

Bond, Z. S. And Sara Garnes (1978) 'Misperceptions of fluent speech.' paper presented at the 14th annual Carnegie Symposium on Cognition: Perception and Production of Fluent Speech, Pittsburgh, PA.

Clarke, Arthur C. Rendezvous with Rama, 1973.

Miller, George A. And Patricia E. Nicely (1955), 'An Analysis of Perceptual Confusions among Some English Consonants,' Jl. Of the Acoustical Society of America 27, 338-352.

Shih, May, Guy Carden, and Linda Lane (1977), 'Unobtrusive Data Collection: an Evaluation of the Appropriate Response Methodology,' paper presented at the Linguistic Society of America, Chicago, IL.

Warmuth, Donald B. (1975) 'Speech Synthesis by a Programmable Digital Filter,' M.S. Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH.

TABLE I. Mean judgements of the intelligibility of the elliptic speech texts.

Condition:	Speech	Speech with noise
Nasality	3.4	4.1
Place	3.7	4.8
Vowels	2.0	2.4
Voicing	5.0	5.8

TABLE II. Number of restorations and veridical reports

Experimental Condition:	Veridical Report	Restoration
Written report		
No noise	10	19
S/N +5 dB	10	32
Shadowing		
No noise	15	32
S/N +5dB	9	47

Figures

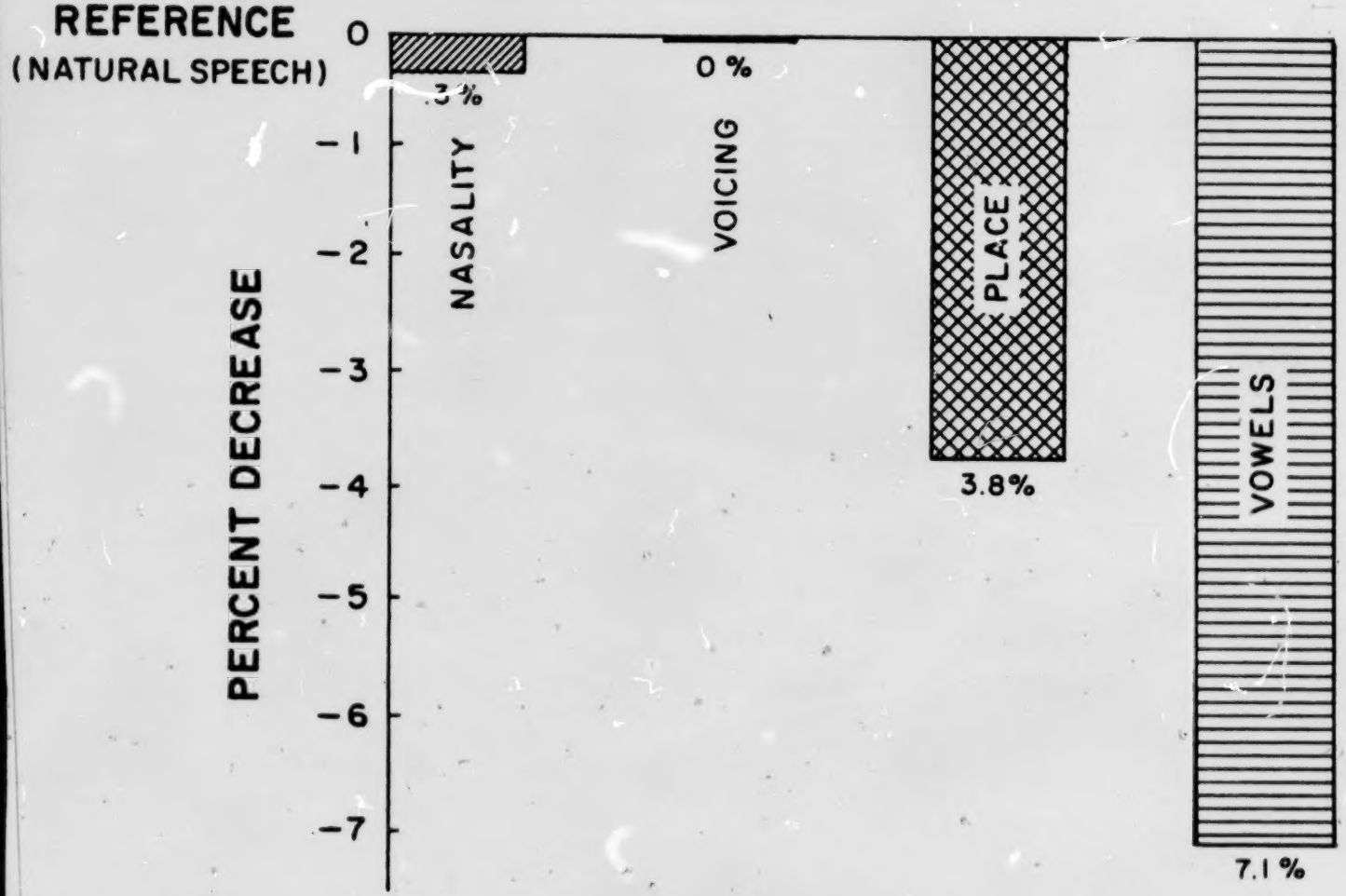
Fig. 1. With reference to natural speech, the per cent decrement in correct report in the memory task is given for the four phonological distortion conditions.

Fig. 2. With reference to natural speech, the per cent decrement in correct report in the memory task is given for the four phonological distortion conditions, when presented under masking noise.

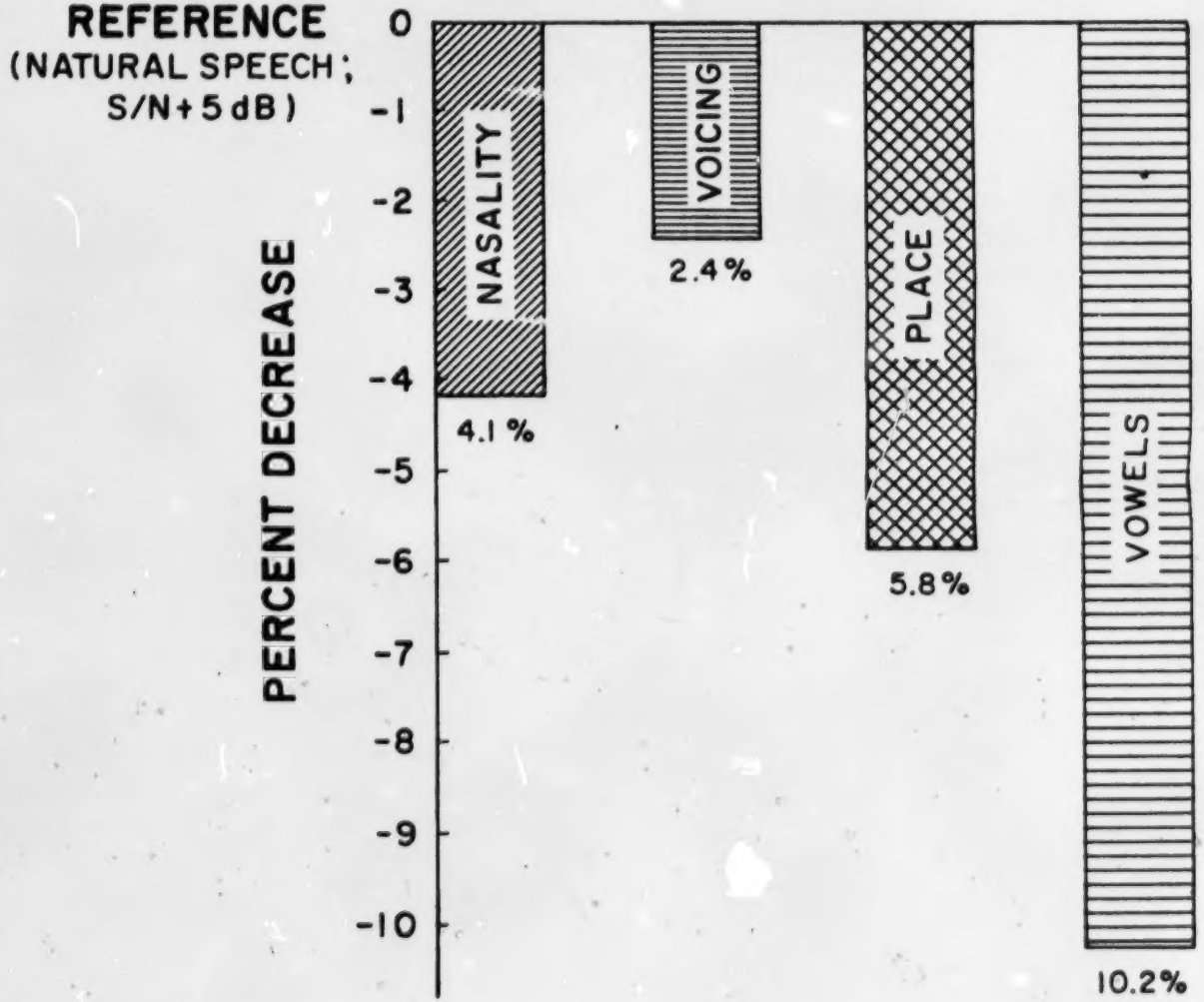
Fig. 3. With reference to natural speech, the per cent decrement in shadowing performance is given for the four phonological distortion conditions.

Fig. 4. With reference to natural speech, the percent decrement in shadowing performance is given for the four phonological distortion conditions, when presented under masking noise.

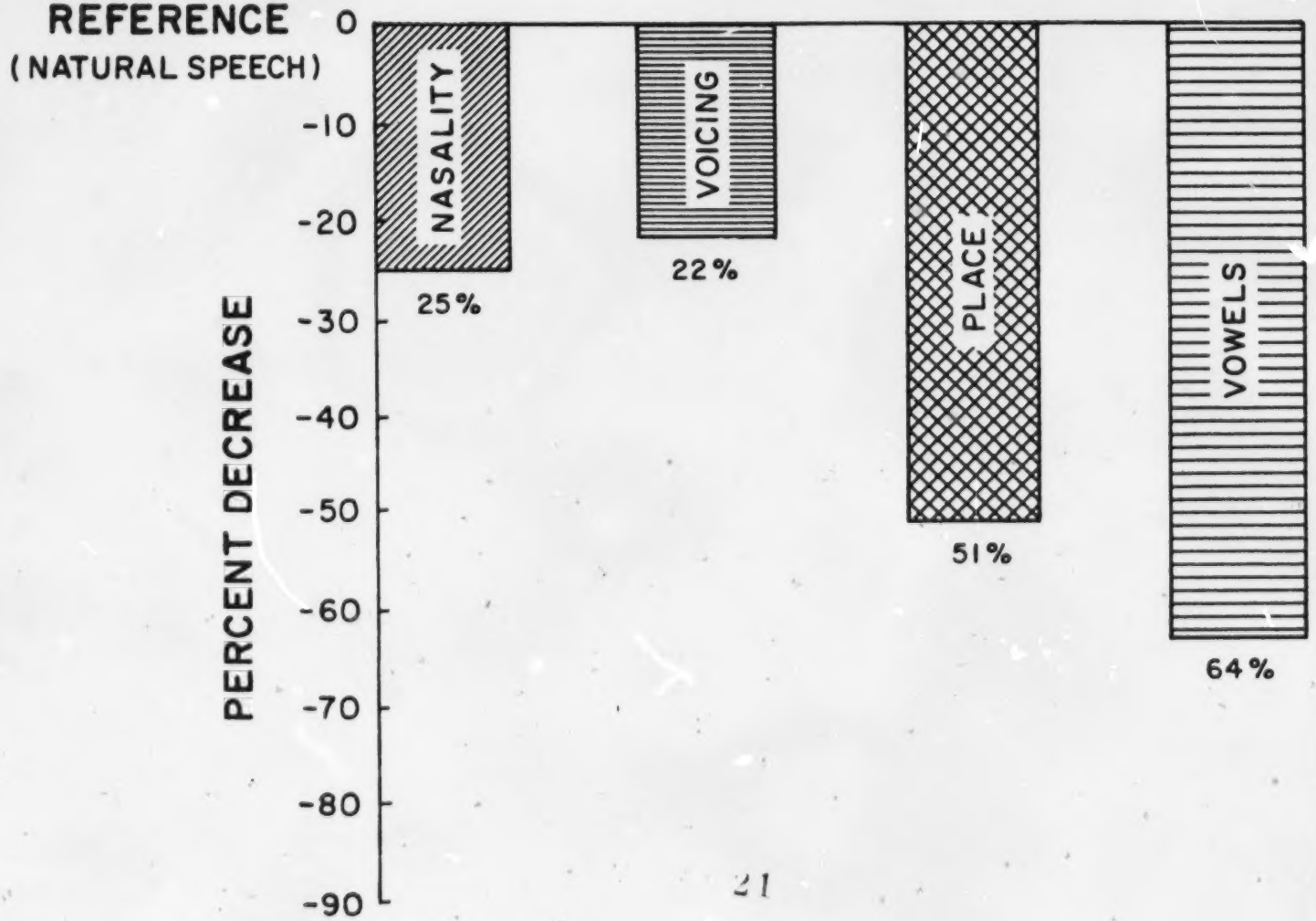
MEMORY FOR PARAGRAPHS: DECREMENT UNDER PHONOLOGICAL DISTORTIONS



MEMORY FOR PARAGRAPHS:
DECREMENT UNDER PHONOLOGICAL DISTORTIONS AND MASKING NOISE (S/N + 5 dB)



SHADOWING PARAGRAPHS: DECREMENT UNDER PHONOLOGICAL DISTORTIONS



SHADOWING PARAGRAPHS:
DECREMENT UNDER PHONOLOGICAL DISTORTIONS AND MASKING NOISE (S/N + 5 dB)